Some basic matrix algebra: Vectors

- $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$: an array of n real numbers, $x_1, \cdots, x_p \in \mathbb{R}$
- Geometrically? directed line in p dimensional space with $i-$th entry x_i along the $i-$th axis.
- Vector operations: for $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$
  - addition: $\mathbf{x} + \mathbf{y}$, $\mathbf{x} - \mathbf{y}$
  - scaler multiplication: $c\mathbf{x}$, for $c \in \mathbb{R}$
  - $\text{span}(\mathbf{x}_1, \dots, \mathbf{x}_k) = \{c_1\mathbf{x}_1 + \cdots + c_k\mathbf{x}_k, c_i \in \mathbb{R}\}$: linear combination of vectors
  - $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ are linearly dependent if $\exists a_1, \cdots, a_k$, not all zero, s.t. $a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k = 0$.
    - what this implies: at least one vector in the set can be written as a linear combination of the other vectors
    - example
  - If the set of vectors are not linearly dependent, we say they are linearly independent.

Vectors

- length of a vector: $L(\mathbf{x}) = \sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$
- angle between the vectors: $\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{x_1y_1 + x_2y_2 + \cdots x_py_p}{L(\mathbf{x})L(\mathbf{y})}$
- inner product $\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \cdots x_py_p$: sum of component products
- relation between angle and inner product: $\cos(\angle(\mathbf{x}, \mathbf{y})) = \frac{\mathbf{x}^T\mathbf{y}}{L(\mathbf{x})L(\mathbf{y})}$
- $\mathbf{x}$ and $\mathbf{y}$ are perpendicular when the angle between the two is $\pi/2$(or $3\pi/2$) and we write $\mathbf{x} \perp \mathbf{y}$
- $\mathbf{x} \perp \mathbf{y}$ if and only if $\mathbf{x}^T\mathbf{y} = 0$.
- projection of a vector $\mathbf{x}$ on a vector $\mathbf{y}$: $\frac{\mathbf{x}^T\mathbf{y}}{L(\mathbf{y})^2}\mathbf{y}$

Vectors

- Euclidean distance between the two vectors:

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + \cdots + (x_p - y_p)^2} = \sqrt{(\mathbf{x} - \mathbf{y})^T(\mathbf{x} - \mathbf{y})}$$

- In general, any distance measure can be used, but it has to satisfy
  - $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$
  - $d(\mathbf{x}, \mathbf{y}) > 0$ if $\mathbf{x} \neq \mathbf{y}$
  - $d(\mathbf{x}, \mathbf{y}) = 0$ if $\mathbf{x} = \mathbf{y}$
  - $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y})$ for any point $\mathbf{z} \in \mathbb{R}^p$.

Matrix

- matrix $\mathbf{A}_{n\times p}$: collection of numbers with n rows and p columns

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{np} \end{bmatrix}$$

  or use $\mathbf{A} = \{a_{ij}\}$

- column of space of $\mathbf{A}$, $C(\mathbf{A}) = \text{ span}(\mathbf{a}_1, \dots, \mathbf{a}_p) = \{c_1\mathbf{a}_1 + \dots + c_p\mathbf{a}_p | c_i \in \mathbb{R}\}$, where $\mathbf{a}_1, \cdots, \mathbf{a}_p$ are the columns of $\mathbf{A}$.

  $C(\mathbf{X}^T) =?$

- matrix operations
  - $\mathbf{A}^T = \{a_{ji}\}$
  - $\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ji}\}$, $\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ji}\}$
  - $c\mathbf{A} = \{ca_{ij}\}$, $c \in R$
  - $\mathbf{A}_{n\times p}\mathbf{B}_{p\times m} = \{\sum_{k=1}^{p} a_{ik}b_{kj}\}$
  - rank$(\mathbf{A})$ = the maximum number of linearly independent rows (columns).
  - trace $(\mathbf{A}) = \sum_{i=1}^{p} a_{ii}$

Square Matrix

- Determinant: $det(\mathbf{A}) = |\mathbf{A}| = \sum(-1)^{|\pi|}a_{1\pi(1)}\cdots a_{p\pi(p)}$, the summation is over all permutations π of $\{1,\ldots,p\}$, and $|\pi| = 0$ if the permutation can be written as a product of an even number of transposition and $|\pi| = 1$ otherwise.

- Inverse: if there exists $\mathbf{B}$ such that $\mathbf{AB} = \mathbf{BB} = \mathbf{I}$, then $\mathbf{B}$ is called the inverse of $\mathbf{A}$ and denoted by $\mathbf{A}^{-1}$.
  - Inverse exists if the k columns of $\mathbf{A}$ are linearly independent.
- Orthogonal Matrices: $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$
  - $\mathbf{q}^T\mathbf{q} = I(i = j)$: length of the column=1 and i and $j-$th columns are orthogonal.

Projection Matrix

- A square matrix $\mathbf{P}$ is called a projection matrix in $\mathbb{R}^p$ if and only if $\mathbf{P} = \mathbf{P}^T = \mathbf{P}^2$ (idempotent).

- For $\mathbf{X}$ $n \times p$ matrix, consider
$$\mathbf{P} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T.$$
Then, the matrix $\mathbf{P}$ projects any vector in $\mathbb{R}^n$ onto the column space $C(\mathbf{X})$ of $\mathbf{X}$.

Eigenvalues and Eigenvectors

- Eigenvalues and eigenvectors: For $\mathbf{A} : p \times p$, $\exists \lambda \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^p$ s.t. $\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$, then we say λ is the eigenvalue and $\mathbf{v}$ is the eigenvector of $\mathbf{A}$.

- Spectral Decomposition: Let $\mathbf{A}$ be a $p \times p$ square symmetric matrix. Then, $\mathbf{A}$ has p pairs of eigenvalues and (normalized) eigenvectors pairs. Specifically, $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^T = \sum_{i=1}^{p} \lambda_i \mathbf{v}_i \mathbf{v}_i^T,$$

  where $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_p]$ is an orthonormal matrix consisting of eigenvectors.

- Let $\lambda_1, \ldots, \lambda_p$ be eigenvalues of $\mathbf{A}$. Then,

$$|\mathbf{A}| = \prod_{i=1}^{p} \lambda_i$$

  and

$$\mathrm{trace}(\mathbf{A}) = \sum_{i=1}^{p} \lambda_i.$$

Quadratic forms

- Study of variance and interrelationship is often based on (squared of) distances.
- Squared distances can be expressed in terms of matrix product called quadratic forms.

Let $\mathbf{A}$ be $p \times p$ square and symmetric matrix. Consider a quadratic form Q:

$$Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} = \sum_{i,j} x_i a_{ij} x_j$$

- $\mathbf{A}$ is said to be non-negative definite ($\mathbf{A} \geq 0$) if $Q(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$, $\forall \mathbf{x} \in \mathbb{R}^p$.
- If $=$ holds only for $\mathbf{x} = \mathbf{0}$, $\mathbf{A}$ is said to be positive definite $\mathbf{A} > 0$.
- $\mathbf{A}$ is positive definite $\Leftrightarrow \lambda_i > 0, \forall i = 1, \dots, p$

  Why?
- For $\mathbf{A} > 0$, let $\mathbf{A} = \mathbf{V} \Lambda \mathbf{V}^T$ be the spectral decomposition of $\mathbf{A}$. Then, $\mathbf{A}^{-1} = \mathbf{V} \Lambda^{-1} \mathbf{V}^T$.
- For $\mathbf{A} > 0$, $\mathbf{V} \Lambda^{1/2} \mathbf{V}^T$ where $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_i})$ is called the square root of $\mathbf{A}$ and denoted by $\mathbf{A}^{1/2}$.